COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 33, No. 4, pp. 851–871, 2004

Data-Driven Discontinuity Detection in Derivatives of a Regression Function

I. Gijbels* and A.-C. Goderniaux

Institut de Statistique, Université catholique de Louvain,
Louvain-la-Neuve, Belgium

ABSTRACT

This paper provides a fully data-driven procedure for estimating the locations of jump discontinuities occuring in the kth derivative of an unknown regression function. The basic ingredients for the procedure are a two-step method for estimating the locations of the jump discontinuities, a bootstrap procedure for selecting the smoothing parameters involved in this estimation, and a cross-validation method for estimating the number of discontinuities in a derivative function. The paper extends ideas developed for change point detection in the regression function itself by Gijbels and Goderniaux [Gijbels, I., Goderniaux, A.-C. (2004). Bandwidth selection for change point estimation in nonparametric regression. *Technometrics* 46:76–86].

*Correspondence: I. Gijbels, Institut de Statistique, Université catholique de Louvain, 20 Voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium; Fax: 32-10-473032; E-mail: gijbels@stat.ucl.ac.be.

DOI: 10.1081/STA-120028730 0361-0926 (Print); 1532-415X (Online)

A simulation study illustrates the performance of the procedure, and applications to some real data demonstrate its use.

Key Words: Abrupt change; Bandwidth; Bootstrap; Cross-validation; Derivative function; Least-squares fitting; Local polynomial regression.

1. INTRODUCTION

Jump discontinuities represent only one type of irregularities that might occur in an otherwise smooth regression function. Other types of irregularities include changes in a derivative of the regression function. An example is a jump discontinuity in the first derivative which would appear as an "abrupt change in the direction" of the regression function itself. A jump discontinuity in the second derivative would appear as a "shoulderpoint" in the regression function itself. Such type of irregularities may show up when estimating regression curves in applications. Consider for example the Motorcycle data, reported on by Schmidt et al. (1981). These are 123 measurements on test objects that underwent a simulated motorcycle collision. Recorded were the time (in milliseconds) after the impact (the variable X) and the head acceleration (in g) of the test object (the variable Y). Looking at the raw data, presented in Fig. 5.5 in Sec. 5, one can "suspect" some changes in the direction of the acceleration somewhere around roughly 15, 23 and 32 milliseconds. In Sec. 5 we will analyze this data set using our data-driven procedure for detecting discontinuities in derivatives of the regression function.

Furthermore, if the interest is in estimating the kth derivative of a regression function, then it is better to obtain an estimator that is adapted to possible jump discontinuities in the derivative function. See Fig. 5.1 in Sec. 5, in which we represent three different regression functions (upper panels) for which a jump discontinuity occurs at the point 0.5 in the first derivative, for Figs. 5.1 (a) and (b), and in the second derivative for Fig. 5.1 (c). The lower panels of Fig. 5.1 depict the corresponding true derivative functions as solid lines (first derivatives in Figs. 5.1 (d) and (e), and the second derivative function in Fig. 5.1 (f)). Presented in these lower panels are also the estimated derivative curves, adapted to the estimated jump discontinuities in the derivative functions (the dotted curves) together with smooth estimates of the derivative curves (the long-dashed curves). The estimates are based on simulated samples of size $n = 100$. Figure 5.1 clearly reveals that the estimates of the derivative functions assuming smooth derivatives are quite different from the estimates of the derivative functions allowing for possible jump points. For more details of these and other simulated examples see Sec. 5.

There are various approaches in the literature dealing with nonparametric regression with abrupt changes in a derivative. Hall and Titterington (1992) proposed a kernel-based estimation method to estimating curves with peaks and edges and Jose and Ismail (1997) rely on the analysis of residuals. Müller (1992) and Wu and Chu (1993), among others, have suggested methods based on differences of nonparametric kernel estimates. Chu (1994) estimated change points in a nonparametric regression function via kernel density estimation using binning. See also Qiu (1994) for kernel based methods when the number of change points is unknown. Eubank and Speckman (1994), Speckman (1994,1995) and Cline et al. (1995) considered semiparametric spline-based methods. Local polynomial procedures have been used by McDonald and Owen (1986), Horváth and Kokoszka (1997), Qiu and Yandell (1998) and Spokoiny (1998), among others. For wavelet-based methods see for example Wang (1995) and Raimondo (1998). All methods have in common that they involve the choice of some kind of smoothing parameters, and the performance of the methods often depends heavily on these choices.

In this paper we provide a fully data-driven procedure for estimating jump discontinuities in a derivative curve. The method also includes a data-driven way of determining the number of discontinuities in a derivative curve. The data-driven procedure developed here is a generalization of the procedure for estimating regression curves with jump discontinuities proposed by Gijbels et al. (1999, 2004) and studied further by Gijbels and Goderniaux (2004). Such a generalization requires an appropriate choice of a diagnostic function and a parametric family for least-squares fitting when dealing with estimation of a (known) number of jump discontinuities. In order to estimate the number of discontinuities we rely on a cross-validation method which basically combines cross-validation ideas discussed in Müller et al. (1987) in the context of bandwidth selection for (smooth) derivative curves and in Müller and Stadtmüller (1999) in the context of estimation of unsmooth regression functions. So, this paper brings together several ideas that have been used in different contexts, and exploits them in the current context of detecting abrupt changes in a regression function and/or its derivatives. It is, to our knowledge, the first paper proposing a fully data-driven method for the studied problem.

The paper is organized as follows. In Sec. 2 we focus on the case of jump discontinuities in the first derivative. We introduce the estimation method and the algorithm to choose the bandwidth parameters.

The generalization to the case of jump discontinuities in a higher order derivative is provided in Sec. 3. In Sec. 4, we discuss a cross-validation criterion to estimate the number of discontinuities appearing in the kth derivative of the regression function. The performance of the data-driven detection procedure is illustrated via a simulation study in Sec. 5. In that section we also present the analysis of the data set described above. In a last section, we provide some discussion.

2. ESTIMATION OF JUMP DISCONTINUITIES IN THE FIRST DERIVATIVE

2.1. Statistical Model

We assume that a sample of n data pairs $\chi = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ is observed, generated from the model

$$Y_i = g(X_i) + \varepsilon_i, \quad 1 \leq i \leq n.$$

We consider design points X_i which are either regularly spaced on $I = [0, 1]$ or are the order statistics of a random sample from a distribution having a density f supported on I. The errors ε_i are assumed to be independent and identically distributed with zero mean and finite variance σ^2. We suppose that $g(\cdot)$, the unknown regression function is such that its first derivative is continuous except at an unknown finite number of jump discontinuities. Denote this unknown number of jump discontinuities by ν.

We consider first the case of a single jump discontinuity, appearing at the location $x_0 \in]0, 1[$ in the first derivative. The generalization to the case of more than one discontinuity in the derivative function is discussed in Sec. 2.4. In Sec. 3 we generalize the method to estimation of the locations of jump discontinuities in the kth derivative of g.

2.2. Estimation Procedure

We adapt the two-step estimation method as discussed by Gijbels et al. (1999, 2004) in the case of jump discontinuities in the regression function itself, to the case of jump discontinuities in the first derivative. This adaptation is straightforward, and consists of two steps: in a first step a preliminary estimator of x_0, the location of the jump discontinuity in the first derivative, is obtained via the evaluation of an appropriate diagnostic function; in a second step an improved estimator of x_0 is obtained by least-squares fitting of an appropriate parametric model in a small interval around the initial estimator of x_0.

2.2.1. Diagnostic Step

A diagnostic function is used to obtain a first estimator $\tilde{x}_0$ of x_0. A way to detect a jump discontinuity in the first derivative is by looking at locations with high second order derivatives. So we suggest to consider the second derivative of a Nadaraya–Watson kernel estimator (see Nadaraya, 1964; Watson, 1964) and define the diagnostic function D by

$$D(x,h_1) = \frac{\partial^2}{\partial x^2}\left(\frac{\sum_{i=1}^n K\{(x-X_i)/h_1\}Y_i}{\sum_{i=1}^n K\{(x-X_i)/h_1\}}\right), \tag{2.1}$$

where K is a compactly supported twice differentiable kernel function and $h_1 > 0$ is a bandwidth. A first rough estimator of x_0 is then given by

$$\tilde{x}_0 = \underset{x\in]vh_1,1-vh_1[}{\operatorname{argmax}} |D(x,h_1)|,$$

where $[-v, v]$ denotes the support of K.

Note that in case of equally-spaced design we can also use the second derivative of the numerator of a Nadaraya–Watson kernel estimator as a diagnostic function:

$$D(x,h_1) = \frac{1}{nh_1^3}\sum_{i=1}^n K^{(2)}\{(x-X_i)/h_1\}Y_i\ ,$$

since this will be proportional to an estimator for the second derivative of the regression function.

2.2.2. Least-Squares Step

The aim of this second step is to improve the initial estimator $\tilde{x}_0$ of x_0. Therefore we construct an interval concentrated around $\tilde{x}_0$, denoted as $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$, with $h_2 > 0$, to which x_0 belongs with high probability. Denote by $\{i_1, i_1+1, \dots, i_2\}$ the set of integers i such that $X_i \in [\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$. Suppose that the change point occurs between the two design points X_{i_0} and X_{i_0+1}. We discuss how to estimate i_0 and hence x_0. To improve the performance of $\tilde{x}_0$, we fit via the least-squares method a linear function (a first order polynomial) to the left and the right of the point X_{i_0} in the interval $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$. More precisely, we search for the value of i_0 that minimizes the sum of squares

$$\sum_{i=i_1}^{i_0}\{Y_i - (a_1 + b_1 X_i)\}^2 + \sum_{i=i_0+1}^{i_2}\{Y_i - (a_2 + b_2 X_i)\}^2 \tag{2.2}$$

where

$$b_j = \frac{\sum_{i=r_j}^{s_j}(X_i - \overline{X}_j)(Y_i - \overline{Y}_j)}{\sum_{i=r_j}^{s_j}(X_i - \overline{X}_j)^2} \quad \text{and} \quad a_j = \overline{Y}_j - b_j \overline{X}_j$$

with $r_1 = i_1, s_1 = i_0$ and $r_2 = i_0 + 1, s_2 = i_2$. The quantities a_1 and b_1, and a_2 and b_2 are nothing but the usual estimated coefficients of a linear regression model.

Denote by $\hat{\imath}_0$ the minimizer of the sum of squares (2.2). The final estimator for x_0 is then defined as the mid-point between $X_{\hat{\imath}_0}$ and $X_{\hat{\imath}_0+1}$:

$$\hat{x}_0 = \frac{1}{2}(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1}).$$

A clear advantage of this two-step method above a one-step method is the following: if the first step results in a bad estimation of the true discontinuity point, the second step can still correct for this. See Gijbels et al. (1999) for illustrations of this important advantage in the regression case. See also Gijbels and Goderniaux (2004) for a finite sample comparison of this two-step method with some one-step procedures in the regression case. The comparison in the latter paper illustrates the advantages of this two-step procedure, and discusses several interesting options in the procedure. Another two-stage procedure for estimating change-points in regression models has been proposed by Müller and Song (1997).

2.3. Choice of the Bandwidth Parameters

The two-step method described above involves two smoothing parameters, the bandwidths h_1 and h_2. The choice of these parameters is rather crucial. We now discuss data-driven choices of these bandwidths.

The diagnostic function depends on the bandwidth h_1, but at a jump discontinuity it will consistently be large for many h_1 values. We then identify the jump discontinuity as the point x in the neighbourhood of which $|D(x, h_1)|$ is consistently large for a range of values of h_1, and take the smallest h_1 value for which this still holds (i.e., decreasing the h_1 value further would introduce artificial peaks at other locations). We consider a set of decreasing h_1 values, namely $h_{1,i} = h_0 r^i$, for $i = 0, 1, 2, \ldots$, with $h_0 > 0$ and $0 < r < 1$. The choice of the biggest h_1 value in this set (i.e., the value h_0) and the choice of the multiplication factor r are not important. Safe choices are h_0 big enough (say up to at most half of the length of the domain of the regression function) and r close to one. See Gijbels and Goderniaux (2004) for more details when dealing with detection of jump discontinuities in the regression function itself.

The choice of the bandwidth h_2 is also crucial since the least-squares fit with a piecewise linear function will be quite bad if the interval $[\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$ is too large such that the unknown function g is far from a piecewise linear function in that interval. Suppose that we have obtained an estimator $\hat{\imath}_0$ (resp. $\hat{x}_0$) of i_0 (resp. x_0) by the two-step method explained in Sec. 2.2, using the data-driven choice of h_1 as explained above and a certain fixed bandwidth h_2. The estimator $\hat{\imath}_0$ is integer-valued and may differ in absolute value from the theoretical (random) i_0 by 0, 1, 2, ... Of course we would like the estimator $\hat{\imath}_0$ to be equal to i_0 with high probability. For choosing the bandwidth h_2 we propose a bootstrap procedure to estimate $P(\hat{\imath}_0 - i_0 = 0)$ for a large set of candidate h_2 values, denoted as $h_{2,j},\ j = 0, \ldots, H$. We then select that bandwidth value h_2 for which the bootstrap estimate of the probability $P(\hat{\imath}_0 - i_0 = 0)$ is largest. The bootstrap algorithm for estimating this probability reads as follows (see also Gijbels et al., 2004).

Step 1. Estimation of g and Computation of Residuals.

Let $\hat{x}_0 = \frac{1}{2}(X_{\hat{\imath}_0} + X_{\hat{\imath}_0+1})$ denote the estimator introduced in Sec. 2.2. Using local linear regression (see for example Fan and Gijbels, 1996), with cross-validation bandwidth selector, we construct $\hat{g}$ on $[0, \hat{x}_0]$ and $[\hat{x}_0, 1]$. We define $\tilde{\varepsilon}_i = Y_i - \hat{g}(X_i)$ for $i = 1, \ldots, n$, and $\bar{\varepsilon}$ the mean of $\tilde{\varepsilon}_i$. Finally, we put $\hat{\varepsilon}_i = \tilde{\varepsilon}_i - \bar{\varepsilon}$, the centralized estimated residuals.

Step 2. Monte Carlo Simulation.

Conditional on the observed sample $\chi = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, we consider $\varepsilon_1^*, \ldots, \varepsilon_n^*$ a resample drawn randomly with replacement from the set $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$. We define

$$Y_i^* = \hat{g}(X_i) + \varepsilon_i^*, \quad i = 1, \ldots, n. \tag{2.3}$$

Then $\chi^* = \{(X_1, Y_1^*), \ldots, (X_n, Y_n^*)\}$ is the bootstrap version of χ.

Step 3. Determination of the Bootstrap Probability.

Using the method described in Sec. 2.2, we compute the analogue $\hat{i}_0^*$ and $\hat{x}_0^* = \frac{1}{2}(X_{\hat{i}_0^*} + X_{\hat{i}_0^*+1})$ of $\hat{i}_0$ and $\hat{x}_0$ for the resample χ^* rather than the sample χ. With B bootstrap replications, we have B values of $\hat{i}_0^*$, denoted by $\hat{i}_0^{*b}, b = 1, 2, \ldots, B$, and we evaluate the discrete probability $P(\hat{i}_0^* - \hat{i}_0 = 0 \mid \chi)$ via

$$\frac{1}{B}\sum_{b=1}^{B} \#\{b : \hat{i}_0^{*b} = \hat{i}_0\}.$$

With this bootstrap algorithm we have a data-driven procedure for estimating the jump discontinuity in the derivative curve: h_1 is automatically chosen as indicated above, and the bandwidth h_2 in the least-squares step is taken to be that bandwidth from the set of possible bandwidths for which the bootstrap estimate of the probability $P(\hat{i}_0 - i_0 = 0)$ is largest.

Note that with this data-driven procedure we opted for allowing for two possibly different bandwidths in the two steps of the estimation method. Alternatively, one could consider taking the same bandwidths in the diagnostic and the least-squares step, and select that single bandwidth via the bootstrap selection procedure described above. Both alternative data-driven procedures have been evaluated via extensive simulation by Gijbels and Goderniaux (2004) in the context of detecting jump discontinuities in the regression function itself. The conclusion was that the more general and more flexible two bandwidths option performs slightly better and hence we opted for this here too.

Note also that for estimating g in the bootstrap algorithm we estimate the jump point of the first derivative, and then estimate via local linear fitting the function g on each of the two intervals separated by the estimated jump point. As always in practice, the estimated curves are calculated for a grid of points, as fine as one wants. For points in between two grid points the value of the estimate is obtained by simple linear interpolation. In our bootstrap implementation we did not force the two local fits on both sides to have the same value at the change point location, since both values are very close in any case (given the continuity of the underlying g). The two local fits could be joined using linear interpolation in order to obtain a continuous estimate on the whole domain (but again, there is no real need for this). This method of estimating separately on the segments determined by the estimated jump point (of the first derivative) is surely not the most efficient way for estimating an unknown regression curve, knowing that the curve shows a jump point in its derivative. See also Sec. 6 for some further discussion.

2.4. The Case of More Than One Jump Discontinuity

The generalization to more than one jump discontinuity is quite straightforward, at least to some extent. Assume that there are ν jump discontinuities in the first derivative. One would then look for the ν local maxima in the diagnostic function $|D(x, h_1)|$ defined in (2.1), and would improve upon initial estimates of the locations by taking a small interval around each initial estimate and fitting via least-squares piecewise linear functions on each interval. Although this seems straightforward, we are not recommending to use this in practice. The reason is that the identification of local maxima corresponding to the jump discontinuities can be somewhat cumbersome. This identification problem can already be an issue when dealing with one single jump discontinuity, and is even more an issue when dealing with two or more jump discontinuities (well separated or not). In short, the identification problem might occur when handling functions g for which the derivative shows steep decreasing or increasing parts, which can blur the detection of jump points. It should be noted that any estimation method will show difficulties with such cases, not only the diagnostic function considered in this paper. For the specific case of our diagnostic function, this would mean that the diagnostic function could achieve its maximum at such points of steep increase or decrease. The solution to the problem is not so difficult though, and consists of a carefully-designed iterative algorithm which tracks back the maximum (or local maxima) associated to a jump discontinuity. For details of this identification problem and the iteration algorithm as a remedy for it, we refer the readers to Gijbels et al. (1999, 2004) or Gijbels and Goderniaux (2004).

In case of more than one jump discontinuity, the above mentioned identification problem can even be more severe, and hence in this case we recommend to use as a default the iterative algorithm which locates the local maxima of the diagnostic function associated to the jump discontinuities.

3. GENERALIZATION FOR DETECTING JUMP DISCONTINUITIES IN HIGHER ORDER DERIVATIVES

We now discuss how to generalize the data-driven procedure for detecting discontinuities in the kth derivative of the regression function. Suppose that the function g is such that its kth derivative is continuous except at a finite number of unknown points. For simplicity we restrict to the case of equally-spaced design points and as a diagnostic function we use the $(k+1)$th derivative of the numerator of a Nadaraya-Watson kernel estimator:

$$D(x, h_1) = \frac{1}{nh_1^{k+2}} \sum_{i=1}^{n} K^{(k+1)}\{(x - X_i)/h_1\} Y_i,$$

where K is a compactly supported ($[-v, v]$) kernel function that is $(k+1)$ times differentiable. For all other design cases it is preferable to work with the $(k+1)$th derivative of the Nadaraya–Watson kernel estimator or any other consistent nonparametric estimator of the unknown regression function. For simplicity we explain the procedure in case of only one discontinuity. Generalizations to more than one jump point are dealt with as indicated in Sec. 2.4 using the appropriate diagnostic function.

An initial estimator of the location x_0 of the jump discontinuity in the kth derivative of the regression function is given by

$$\tilde{x}_0 = \operatorname*{argmax}_{x \in]vh_1, 1-vh_1[} |D(x, h_1)|.$$

This rough estimator is then refined by considering a small interval around this initial estimator and fitting via the least-squares method a piecewise kth order polynomial in this interval. More precisely, assuming that the jump point x_0 falls between X_{i_0} and X_{i_0+1}, and denoting by $\{i_1, i_1+1, \ldots, i_2\}$ the set of all indices i for which $X_i \in [\tilde{x}_0 - h_2, \tilde{x}_0 + h_2]$, then we estimate i_0 by $\hat{\imath}_0$ the minimizer of the following sum of squares:

$$\sum_{i=i_1}^{i_0} \left\{ Y_i - \sum_{j=0}^{k} a_j X_i^j \right\}^2 + \sum_{i=i_0+1}^{i_2} \left\{ Y_i - \sum_{j=0}^{k} b_j X_i^j \right\}^2.$$

The coefficients $a_j, b_j, j = 0, \ldots, k$ are the usual coefficients from a global least-squares fit with a polynomial of order k, and are given by

$$a = (a_0, \ldots, a_k)^T = (X^T X)^{-1} X^T Y \quad \text{and}$$
$$b = (b_0, \ldots, b_k)^T = (X'^T X')^{-1} X'^T Y',$$

where the superscript T denotes the transposed of a vector or matrix, and with

$$X = \begin{pmatrix} 1 & X_{i_1} & X_{i_1}^2 & \dots & X_{i_1}^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{i_0} & X_{i_0}^2 & \dots & X_{i_0}^k \end{pmatrix} \qquad Y = \begin{pmatrix} Y_{i_1} \\ \vdots \\ Y_{i_0} \end{pmatrix}$$

and

$$X' = \begin{pmatrix} 1 & X_{i_0+1} & X_{i_0+1}^2 & \dots & X_{i_0+1}^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{i_2} & X_{i_2}^2 & \dots & X_{i_2}^k \end{pmatrix} \qquad Y' = \begin{pmatrix} Y_{i_0+1} \\ \vdots \\ Y_{i_2} \end{pmatrix}.$$

The final estimator of x_0 is then defined as $\hat{x}_0 = \frac{1}{2}(X_{\hat{i}_0} + X_{\hat{i}_0+1})$.

To choose the two different bandwidths involved in this two-step estimation method we use the algorithm described in Sec. 2.3.

4. ESTIMATING THE NUMBER OF DISCONTINUITIES IN A DERIVATIVE

An important issue in discontinuity detection is to determine the number of discontinuities, which is often unknown in practice. Müller and Stadtmüller (1999) propose a cross-validation criterion to estimate the number of discontinuities that appear in the regression function itself. In this section we generalize this method to estimate the number of discontinuities appearing in the kth derivatives of the regression function.

For ease of comprehension we first explain the method when the jump points occur in the first derivative (i.e., $k = 1$). In this case it seems natural to define the cross-validation quantity as follows:

$$\mathrm{CV}^{(1)}(\nu) = \frac{1}{n-1}\sum_{i=1}^{n-1}\left\{\frac{Y_{i+1} - Y_i}{X_{i+1} - X_i} - \hat{g}^{(1)}_{-(i,i+1),\nu}(X_i^{(1)})\right\}^2$$

where ν represents the number of discontinuities, $X_i^{(1)} = (X_{i+1} + X_i)/2$ and $\hat{g}^{(1)}_{-(i,i+1),\nu}(X_i^{(1)})$ is the leave-(2)-out kernel estimator of the first derivative of g, the regression function, based on the data $(X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), (X_{i+2}, Y_{i+2}), \dots, (X_n, Y_n)$ and adapted to the ν estimated jump points of the derivative function. This derivative estimator is obtained by carrying out local polynomial fitting of order 2 on all $\nu + 1$

intervals separated by the estimated jump points, using cross-validated bandwidths (adapted to estimation of the derivative function, see Müller et al., 1987). Note that $(Y_{i+1} - Y_i)/(X_{i+1} - X_i)$ represents the slope of the line between Y_{i+1} and Y_i. We compare this slope with an estimator of the first derivative of the regression function at the point $(X_{i+1} + X_i)/2$. It is clear that we want the cross-validation quantity to be as small as possible.

This method generalizes easily to the case of jump points occuring in the kth derivative of the regression function. Define $X_i^{(0)} = X_i$ and $\delta_i^{(0)} = Y_i$ for $i = 1, \ldots, n$ and put

$$X_i^{(k)} = \frac{1}{2}\left(X_{i+1}^{(k-1)} + X_i^{(k-1)}\right), \quad \delta_i^{(k)} = \frac{\delta_{i+1}^{(k-1)} - \delta_i^{(k-1)}}{X_{i+1}^{(k-1)} - X_i^{(k-1)}}$$

for $i = 1, \ldots, n-k$. The proposed generalization of the above cross-validation quantity is then

$$\mathrm{CV}^{(k)}(\nu) = \frac{1}{n-k}\sum_{i=1}^{n-k}\left(\delta_i^{(k)} - \hat{g}^{(k)}_{-(i,i+k),\nu}\left(X_i^{(k)}\right)\right)^2,$$

where $\hat{g}^{(k)}_{-(i,i+k),\nu}(X_i^{(k)})$ is the leave-$(k+1)$-out kernel estimator of the kth derivative function based on the data $(X_1, Y_1), \ldots, (X_{i-1}, Y_{i-1})$, $(X_{i+k+1}, Y_{i+k+1}), \ldots, (X_n, Y_n)$ and obtained via local polynomial approximation of order $(k+1)$ on each of the $\nu + 1$ intervals defined by the ν estimated jump points of the kth derivative function, and using cross-validation bandwidth selectors adapted to the estimation of derivative curves.

Such a type of cross-validation quantity has been proposed by Müller et al. (1987) in the context of bandwidth selection for estimating the kth derivative of a regression function.

So in practice we calculate the cross-validation quantity for each pre-specified number of discontinuities and then choose that number (of discontinuities) which corresponds with the smallest cross-validation value. To estimate the location of the pre-specified ν discontinuities, $\nu = 0, 1, 2, \ldots$ we propose to use the fully data-driven bootstrap procedure adapted to the derivative case. Finally we estimate the number of discontinuities appearing in the kth derivative by

$$\hat{\nu} = \operatorname*{argmin}_{\nu \in \{0,1,\ldots\}} \mathrm{CV}^{(k)}(\nu).$$

Note that this data-driven procedure is developed for detecting jump points in a derivative function of pre-specified order. See Sec. 6 for a brief discussion on discontinuities appearing in several derivatives.

5. SIMULATION STUDIES AND APPLICATION

5.1. Simulation Study

In this section we evaluate via a simulation study the data-driven estimation procedure. We consider the regression functions

$$g_1(x) = \begin{cases} 2x+1 & \text{if } x \in [0, 0.5] \\ -2x+3 & \text{if } x \in]0.5, 1] \end{cases}$$

$$g_2(x) = \begin{cases} 10x^2 & \text{if } x \in [0, 0.5] \\ -20/7\, x^3 + 20/7 & \text{if } x \in]0.5, 1] \end{cases}$$

$$g_3(x) = \begin{cases} 12x^2 - 2x + 2 & \text{if } x \in [0, 0.5] \\ -12x^2 + 22x - 4 & \text{if } x \in]0.5, 1], \end{cases}$$

and work with fixed equidistant design, $x_i = i/n$ for $i = 1, \ldots, n$. The errors ε_i were taken to be Gaussian with variances $\sigma^2 = 0.1$ or 0.5. We present simulation results for sample sizes $n = 100$ or 200. In the upper panels of Fig. 5.1 we depict the true regression functions g_1, g_2 and g_3 with typical simulated data sets for sample size $n = 100$ and $\sigma^2 = 0.01$. Note that both functions g_1 and g_2 have a single jump discontinuity at $x_0 = 0.5$ appearing in the first derivative. The size of the jump is -4 for g_1 and $-85/7$ for g_2. The function g_3 presents a single jump discontinuity of size -48 at the point $x_0 = 0.5$ in the second derivative. In the lower panels of Fig. 5.1 we present the true derivative functions $g_1^{(1)}, g_2^{(1)}$ and $g_3^{(2)}$ as solid curves along with smooth local quadratic (respectively cubic) estimators as long-dashed curves and estimators adapted to the estimated jump points as dotted curves. The adapted estimates were obtained by local quadratic fitting for the functions g_1 and g_2 and by local cubic fitting for the function g_3 on the two intervals separated by the estimated jump point. Figure 5.1 is based on simulations from an error with relatively small variance. This small error variance is only considered for producing Fig. 5.1 (especially focusing on (d)–(f)), in order to obtain estimated derivative curves that present nice visually. Recall that estimation of derivatives curves is more difficult than estimation of the regression function itself. For all other simulations, focusing on the estimation of the jump points, we consider larger error variances. Note, from Fig. 5.1 (a)–(c), that even with such a small error variance, it is hard to tell from the data what is happening (smooth or non-smooth) with the first (second) derivative.

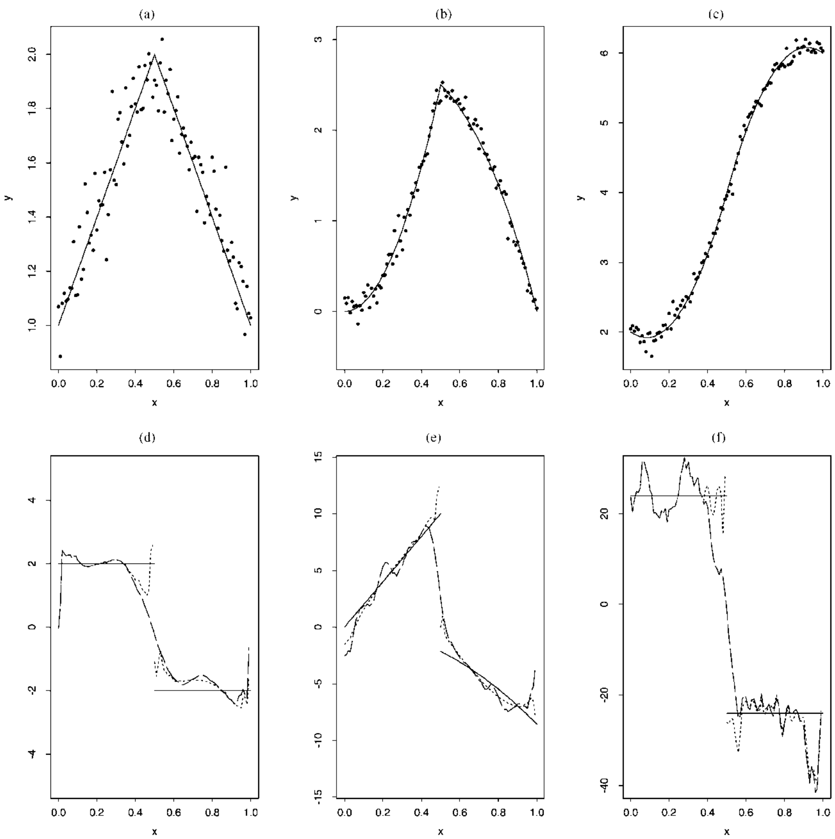

Figure 5.1. Upper panels: The true regression functions (solid curves) with a typical simulated data set of size $n = 100$ and variance $\sigma^2 = 0.01$: Regression functions (a) g_1; (b) g_2; and (c) g_3. Lower panels: the true kth derivative (solid curve) for the three functions along with a smooth local quadratic (for (d) and (e)) and local cubic (for (f)) estimator (long-dashed curve) and an unsmooth estimator (dotted curve). Regression function (d) g_1, $k = 1$; (e) g_2, $k = 1$; and (f) g_3, $k = 2$.

5.1.1. Estimation of the Localisation of the Jump

For the diagnostic function in (2.1), we used a standard Gaussian kernel. In all simulation studies we considered 1000 simulations and the number of bootstrap replicates was $B = 2000$. For determining h_1, the smoothing parameter of the diagnostic function, we search over the

Table 5.1. Simulation results for the functions g_1 and g_2 having one discontinuity in the first derivative at $x_0 = 0.5$, and for the function g_3 having one discontinuity in the second derivative at $x_0 = 0.5$.

		$\sigma^2 = 0.1$		$\sigma^2 = 0.5$	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$
g_1	Mean of $\hat{x}_0$	0.503200	0.501500	0.503630	0.502155
	SD of $\hat{x}_0$	0.041326	0.038776	0.084613	0.068152
g_2	Mean of $\hat{x}_0$	0.496810	0.499380	0.491730	0.488905
	SD of $\hat{x}_0$	0.017943	0.014585	0.041763	0.035615
g_3	Mean of $\hat{x}_0$	0.533650	0.495300	0.560300	0.509960
	SD of $\hat{x}_0$	0.073594	0.047186	0.121593	0.088881

set of bandwidths $h_{1,i} = h_0 r^i$, for $i = 0, 1, 2, \ldots$, with $r = 0.9$ and $h_0 = 0.2$ (respectively $h_0 = 0.1$) when we study the functions g_1 and g_3 (respectively g_2). For the set of potential bandwidths $h_{2,j}, j = 0, \ldots, H$, in the least-squares step we took $h_{2,j} = 0.03 + 0.015j$ for $j = 0, \ldots, 7$.

In Table 5.1 we summarize the simulation results for the functions g_1, g_2 and g_3. Presented are the means and standard deviations of $\hat{x}_0$ across the 1000 simulations. Figure 5.2 presents boxplots of the 1000 estimated values of x_0 for all three functions. If we compare the results of the two functions that present a jump in the first derivative (g_1 and g_2) we can see that the results for the function g_1 are not as good as those for the function g_2. This is related to the fact that the size of the jump in the derivative function is smaller for the function g_1.

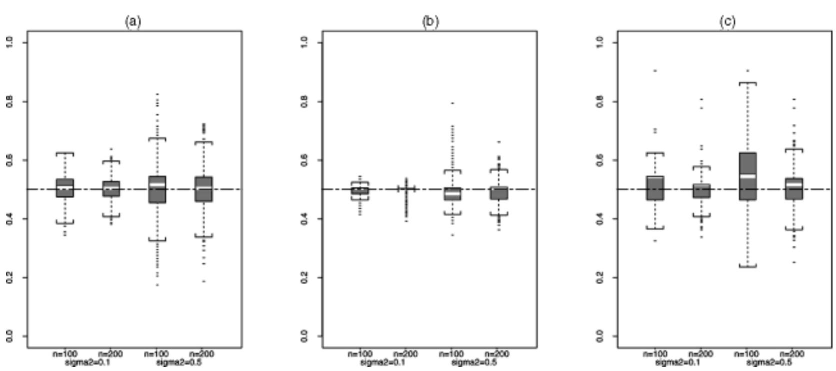

Figure 5.2. Boxplots of the 1000 estimated values $\hat{x}_0$ for x_0 for the functions (a): g_1, (b): g_2 and (c) g_3, for sample sizes $n = 100$ and 200 and $\sigma^2 = 0.1$ and 0.5.

5.1.2. Estimation of the Number of Discontinuities

For each of the functions g_1, g_2 and g_3 we also applied the cross-validation method for determining the number of discontinuities. Here we used 100 simulations and $B = 1000$ bootstrap samples. Table 5.2 summarizes the simulation results for the three functions for sample sizes $n = 100$ and $n = 250$ and σ^2 equal to 0.1 and 0.5. Presented are the frequencies (out of 100) that the estimated values $\hat{\nu}$ correspond to the specified values. As an illustration we show in Fig. 5.3 the cross-validation quantity as function of ν for a simulated data set of size $n = 100$ generated from the function g_1 with $\sigma^2 = 0.1$. We conclude that for this sample the number of discontinuities should be taken to be one. From Table 5.2 we see that the estimated number of discontinuities is positively biased in the three examples. In fact overestimation of the number of discontinuities is less harmful than underestimation of this number. Indeed, if a point is falsely selected as being a location where the derivative function is discontinuous, then as a consequence the resulting derivative estimate might be discontinuous at that point too. If, on the other hand a discontinuity point would not be detected, then this would result in a far too smooth behaviour of the estimated derivative function in the neighbourhood of such a point.

Table 5.2. Simulation results for the cross-validation choice of ν.

			0	1	2	3	4
g_1	$n = 100$	$\sigma^2 = 0.1$	0	78	12	10	0
One discontinuity in		$\sigma^2 = 0.5$	0	73	16	10	1
the first derivative	$n = 250$	$\sigma^2 = 0.1$	0	83	8	9	0
		$\sigma^2 = 0.5$	0	79	11	9	0
g_2	$n = 100$	$\sigma^2 = 0.1$	1	76	15	6	2
One discontinuity in		$\sigma^2 = 0.5$	1	73	17	8	1
the first derivative	$n = 250$	$\sigma^2 = 0.1$	0	86	4	9	1
		$\sigma^2 = 0.5$	0	83	5	12	0
g_3	$n = 100$	$\sigma^2 = 0.1$	2	64	22	8	4
One discontinuity in		$\sigma^2 = 0.5$	2	61	26	5	7
the second derivative	$n = 250$	$\sigma^2 = 0.1$	0	76	7	12	5
		$\sigma^2 = 0.5$	1	70	23	2	4

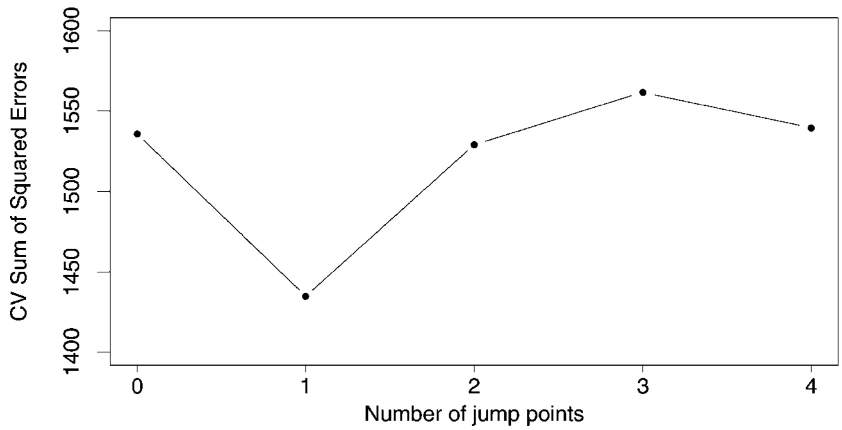

Figure 5.3. Cross-validation sum of squares as function of the number of jump points for a simulated data set of size $n = 100$ from the function g_1 with $\sigma^2 = 0.1$.

5.2. Application

For the Motorcycle data we applied the data-driven method to search for jump discontinuities in the first derivative of the regression function, since we aim at finding changes in direction in the acceleration. For the smoothing parameter h_1 in the diagnostic step we searched over the set $h_{1,i} = h_0 r^i$ with $h_0 = 7$ and $r = 0.9$. For the interval length in the least-squares step we used the set of potential bandwidths $h_{2,j} = 3.5 + 1.0j$ for $j = 0, 1, \ldots, 8$. The cross-validation quantity as function of ν is represented in Fig. 5.4, and from this we fix the number of discontinuities to be three. The data-driven estimation method provided the estimations 14.2, 24.1 and 32.4 for the three change points. These results agree with those obtained by Speckman (1995) who uses a semi-parametric change point method to identify the number and the locations of the change points. Figure 5.5 shows the data along with an adaptive local linear fit.

6. DISCUSSION

In the proposed procedure the estimation of the regression function itself was done as follows. The ν estimated jump discontinuities in the kth derivative define $\nu + 1$ intervals and on each of these intervals we obtain the local linear fit. These fits are then joined together to get the global

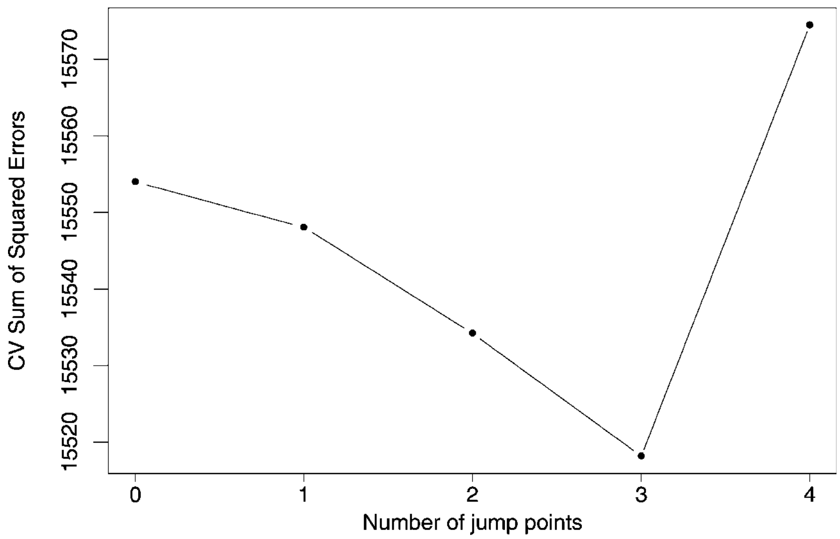

Figure 5.4. Cross-validation sum of squares as function of the number of jump points for the motorcycle data.

estimator for the regression function. This estimation method does not use the fact that the jump discontinuities occur in the kth derivative of the function. One way to use more directly the information about jump discontinuities in the kth derivative is to fit splines of order k with knots

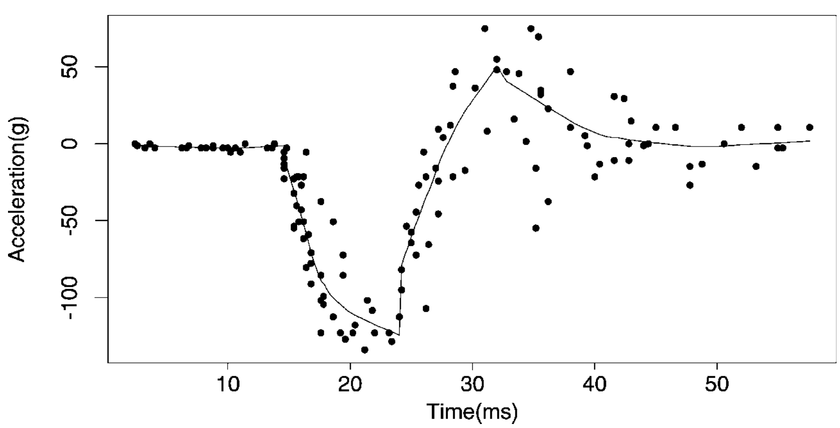

Figure 5.5. The motorcycle data (points) with a local linear fit adapted to change-points.

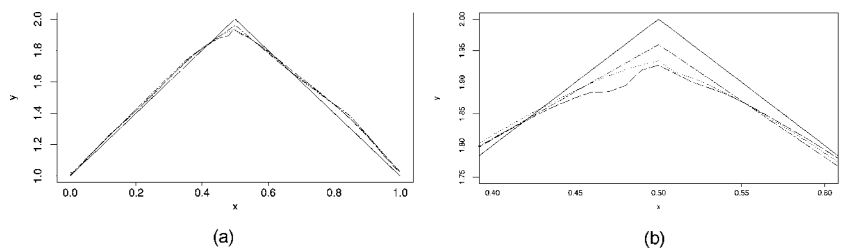

Figure 6.1. The true regression function g_1 (solid curve) with the local linear fits joined together (dotted curve), the estimated curve obtained by integration of the estimate of the first derivative (long-dashed curve) and finally the linear spline fit (dotted-dashed curve). The right panel shows a zoom-in of the left panel.

at the estimated ν jump discontinuities. In this way, the resulting estimator would have continuous first $k-1$ derivatives. Another way to include the information about the non-smoothness of the kth derivative in the estimation of the regression function itself is to estimate the kth derivative by a local polynomial fit of degree $k+1$, adapted to change points and to integrate it k times to reconstruct the function itself.

In order to appreciate the quality of the different proposed approaches to estimate the regression function with a jump in one derivative, we present in Fig. 6.1 the true regression function g_1 in solid line with the local linear fits joined together (dotted curve), the estimated curve obtained by integration of the estimate of the first derivative (long-dashed curve) and finally the linear spline fit (dotted-dashed curve). The right panel presents a zoom-in of the left panel and shows more in detail what happens around the change-point. We conclude that there are little differences between the three proposed estimated curves. Nevertheless the spline fit performs slightly better.

ACKNOWLEDGMENTS

The authors are grateful to an anonymous referee for careful reading of the first version of the manuscript. Financial support from the contract 'Projet d'Actions de Recherche Concertées' nr 98/03–217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

REFERENCES

Cline, D. B. H., Eubank, R., Speckman, P. (1995). Nonparametric estimation of regression curves with discontinuous derivatives. *J. Stat. Res.* 29:17–30.

Chu, C. K. (1994). Estimation of change-points in a nonparametric regression function through kernel density estimation. *Commun. Stat., Part A–Theory and Methods* 23:3037–3062.

Eubank, R. L., Speckman, P. L. (1994). Nonparametric estimation of functions with jump discontinuities. In: Carlstein, E., Müller, H.-G., Siegmund, D., eds. *Change-Point Problems*. IMS-Lecture Notes. Vol. 23, pp. 130–143.

Fan, J., Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. New York: Chapman and Hall.

Gijbels, I., Goderniaux, A.-C. (2004). Bandwidth selection for change point estimation in nonparametric regression. *Technometrics* 46:76–86.

Gijbels, I., Hall, P., Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Stat. Math.* 51:231–251.

Gijbels, I., Hall, P., Kneip, A. (2004). Interval and band estimation for curves with jumps. *J. App. Probability*. To appear.

Hall, P., Titterington, D. M. (1992). Edge preserving and peak-preserving smoothing. *Technometrics* 34:429–440.

Horváth, L., Kokoszka, P. (1997). Change-point detection with nonparametric regression. *Technical Report,* University of Liverpool.

Jose, C. T., Ismail, B. (1997). Estimation of jump points in nonparametric regression through residual analysis. *Commun. Stat., Part A–Theory and Methods* 26:2583–2607.

McDonald, J. A., Owen, A. B. (1986). Smoothing with split linear fits. *Technometrics* 28:195–208.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Stat.* 20:737–761.

Müller, H.-G., Song, K. -S. (1997). Two-stage change-point estimators in smooth regression models. *Stat. Probability Lett.* 34:323–335.

Müller, H.-G., Stadtmüller, U. (1999). Discontinuous versus smooth regression. *Ann. Stat.* 27:299–337.

Müller, H.-G., Stadtmüller, U., Schmitt, T. (1987). Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika* 74:743–749.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probability Appl.* 9:141–142.

Qiu, P. (1994). Estimation of the number of jumps of the jump regression functions. *Commun. Stat., Theory and Methods* 23:2141–2155.

Qiu, P., Yandell, B. (1998). A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics* 40:141–152.

Raimondo, M. (1998). Minimax estimation of sharp change points. *Ann. Stat.* 26:1379–1397.

Schmidt, G., Mattern, R., Schüler, F. (1981). Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebral column with and without protective helmet under effects of impact. EEC Research Program on Biomechanics of Impacts. Final Report Phase III, Project 65, Institut für Rechtsmedizin, Universität Heidelberg, Germany.

Speckman, P. L. (1994). Detection of change-points in nonparametric regression. Unpublished manuscript.

Speckman, P. L. (1995). Fitting curves with features: Semiparametric change-point methods. *Comput Sci. Stat.* 26:257–264.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Stat.* 26:1356–1378.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* 82:385–397.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Series A* 26:359–372.

Wu, J. S., Chu, C. K. (1993). Modification for boundary effects and jump points in nonparametric regression. *J. Nonparametric Stat.* 2:341–354.